A Visual Sensing Platform for Creating A Smarter Multi-Modal Marine Monitoring Network

Dian Zhang
CLARITY, Dublin City University, Ireland
dian.zhang2@mail.dcu.ie

Edel O'Connor
CLARITY, Dublin City University, Ireland
edel.oconnor@dcu.ie

Kevin McGuinness
CLARITY, Dublin City University, Ireland
kevin.mcguinness@dcu.ie

Noel E. O'Connor
CLARITY, Dublin City University, Ireland
noel.oconnor@dcu.ie

Fiona Regan
MESTECH, Dublin City University, Ireland
fiona.regan@dcu.ie

Alan Smeaton
CLARITY, Dublin City University, Ireland
alan.smeaton@dcu.ie

ABSTRACT

Demands from various scientific and management communities along with legislative requirements at national and international levels have led to a need for innovative research into large-scale, low-cost, reliable monitoring of our marine and freshwater environments. In this paper we demonstrate the benefits of a multi-modal approach to monitoring and how an in-situ sensor network can be enhanced with the use of contextual image data. We provide an outline of the deployment of a visual sensing system at a busy port and the need for monitoring shipping traffic at the port. Subsequently we present an approach for detecting ships in a challenging image dataset and discuss how this can help to create an intelligent marine monitoring network.

Categories and Subject Descriptors

I.4 [**IMAGE PROCESSING AND COMPUTER VISION**]: Miscellaneous; I.5.4 [**Applications**]: Computer vision

Keywords

environmental monitoring, multi-modal sensing, visual sensing

1. INTRODUCTION

Increasing demands from various scientific and management communities for monitoring issues such as climate change, water quality, coastal erosion, flooding and ecosystem change along with legislative requirements at both national and international levels have led to a need for innovative research into large-scale, reliable and sustainable monitoring of our marine and freshwater environments. Coastal and freshwater environments in particular represent vital environmental assets on many levels and need continuous monitoring and protection. Accurately monitoring the quality of these waters can prove very difficult since the associated environmental processes often demonstrate high frequency spatial and temporal variation and are extremely heterogeneous. New technologies are emerging to enable remote autonomous sensing of water systems and subsequently meet the demands for high temporal and spatial monitoring. In our work we extend our conventional understanding of a sensor network or a community of sensor nodes to include diverse data sources and multiple sensing modalities in order to create a smarter marine monitoring network. In particular we focus on the use of visual sensors to complement and enhance the use of an in-situ environmental monitoring network.

In our previous work, we have shown how rainfall radar, a deployed camera and in-situ sensing can be collaboratively combined to optimise a coastal monitoring network [7, 8]. In this work our focus moves to a different site with very different characteristics, dynamics and issues. Poolbeg Marina is located on the lower part of the estuary of the River Liffey, in Dublin, Ireland. Due to the large amount of activity at the site and its importance from an environmental and ecological perspective, the site was equipped with a multi-parameter in-situ sensor, equipped with turbidity, dissolved oxygen, temperature, conductivity and depth probes along with a visual sensing system. The visual sensor continuously sends images back to a cloud server at a relatively low frame rate (approximately 1 frame every 10 seconds). Shipping traffic at the port greatly effects the aquatic ecosystem. Emissions from the large amount of traffic, along with propeller contact, noise, movement and turbulence from the propulsion systems can have multiple effects on the ecosystem including increased turbidity. There are many negative impacts of increased turbidity on the ecosystem and these are well defined in the literature [6]. Analysis of the sensor data demonstrates that ships entering the port often coincide with spikes in data from the turbidity sensor (Figure 1). The same effects are not seen with the activity of small boats in the area. In the following we describe an approach for accurately detecting ships in a challenging image dataset. We achieve very high accuracy which leads to the possibility of smarter marine monitoring via low-cost intelligence provided by contextual information from a visual sensor at the site.

Figure 1: Turbidity readings showing the relationship between tubidity spikes and presence of ship

2. RELATED WORK

Automated analysis of image data for analysing and detecting environmental processes and events has been studied in a wide variety of contexts. For example in previous studies, video systems have been identified as effective tools for coastal monitoring. A prime example of this is a major European research project entitled CoastView [3]. This focused on the development of video systems in support of coastal zone management utilizing Argus technology. Other studies have investigated the use of cameras and analysis of the resulting image data for other forms of environmental monitoring applications. In [4] Graham et al. investigate the use of cameras in determining the dynamics of expanding leaf area for Rhododendron occidentale. Richardson et al. [10] explored whether digital webcam images could be used to monitor spring green-up in a deciduous northern hardwood forest. They concluded that webcams offer an inexpensive means by which phenological changes can be quantified.

3. METHODOLOGY

Usually a continuous monitoring system is installed at a fixed location and does not move over time. This provides an opportunity to model the information of the scene – a background model. Many background modelling techniques have been developed and widely used for detecting moving objects in videos, some of these are reviewed in [9]. However, the unreliable low frame rate of the remote visual sensor unit prevents the use of most of these techniques such as object tracking [5] or optical flow [1] techniques for detection of ships at the scene. Thus we investigate two alternative background modelling techniques that may be able to overcome these issues: Eigenbackgrounds [12] and Gaussian Mixture Models(GMM) [11].

3.1 Eigenbackgrounds

Eigenbackgrounds are formed by calculating the mean and its covariance for a set of images. This covariance matrix is then diagonalized using eigenvalue decomposition. To reduce the dimensionality of the space only the first n largest eigenvectors and associated eigenvalues are kept. For a particular scenario, the optimised number of dimensions n can be defined by calculating the explained variance ratio (EVR) of all components. The EVR is the percentage of variance explained by each of the selected components. A new incoming image, can be projected onto the Eigenspace then back-projected onto the image space. Since the Eigenspace only stores the information of training data, the back-projected image will not contain any objects that do not appear in the training set. Also, the Eigenvalue represents the characteristics of the value distribution of a set of pixels, so it can be used to build a compact and accurate background model of image data with a low and variant frame rate. Figure 2 shows an example of background subtraction using Eigenbackgrounds.

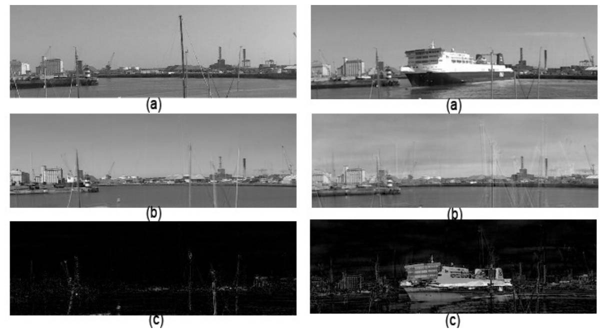

Figure 2: (a) input image, (b) reconstructed image, (c) difference between input image and reconstructed image

3.2 Gaussian Mixture Model Background

In Gaussian Mixture Modelling the history of a particular pixel's values x_0, y_0 is defined as a time series:

$$\{X_1, ..., X_t\} = \{I\left(x_0, y_0, i\right) : 1 \leq i \leq t\} \tag{1}$$

Where I refers to the image sequence and x_i is a vector referring to the intensity value of the pixel. According to [11], the probability of observing the current pixel value is:

$$P\left(X_t\right) = \sum_{i=1}^{K} \omega_{i,t}, \mathcal{N}\left(X_t; \mu_{i,t}, \sigma_{i,t}\right) \tag{2}$$

Where K is the number of distributions. Stauffer and Grimson [11] explain that K is determined by the available memory and computational power. Once the background model is built, for each new incoming image, the probability of each pixel can be calculated. To maximise the difference between foreground and background, the negative log probability is used. Figure 3 shows an example of background subtraction using GMM.

Figure 3: (a) input image, (b) foreground

3.3 Feature Extraction using Histograms of Local Background Residuals (HLBR)

Since our objective is to be able to detect when large objects like large boats and ships enter the harbour, we need to extract a set of features that are sufficiently discriminative to allow us to classify such events. The background subtraction techniques described in the previous section produce a residual image R in which the value of each pixel represents the magnitude of the difference between the observed scene and the background model. If the background model effectively captures the true variation expected in scenes containing normal events, then we expect that the residual R_{ij} will be relatively high in the vicinity of previously unseen objects, and relatively low elsewhere. If, on the other hand, the image contains normal events, we expect that the residual will be comprised primarily of noise and therefore relatively uniform and of low magnitude over all locations (i, j). The above reasoning suggests that global statistics of local residual magnitude may be an effective way to distinguish between images with and without objects of interest. We propose the following Monte Carlo-based approach for generating a global descriptor that captures local variation in the image.

Given an $m \times n$ residual image R, sample K pairs $(x, y)_k$ of random integers such that $x \sim \mathcal{U}(0, m - w)$ and $y \sim \mathcal{U}(0, n - w)$, where $\mathcal{U}(a, b)$ is a uniform distribution over the integers in the half-open interval $[a, b)$, and w is a window size parameter. For each pair $(x, y)_k$ compute the expectation on the residual over the corresponding window as:

$$E[R_{ij}|(x, y)_k] = \frac{1}{w^2} \sum_{i=x}^{x+w} \sum_{j=y}^{y+w} R_{ij}. \tag{3}$$

These expectations capture the local residual magnitude over fixed sized windows in the image. In particular, $E[R_{ij}|(x, y)_k]$ will be relatively large in the vicinity of objects not captured by our background model, and small elsewhere. The global distribution of $E[R_{ij}]$ can, therefore, be used to characterize the local residual magnitude in the image. We used a simple normalized histogram with fixed width bins to capture the distribution of local residual magnitudes produced by (3) to generate a global fixed length descriptor. For our experiments, we used a histogram of 60 bins, producing a length 60 descriptor. The choice of the number of bins and the width parameter w are described in more detail in the experiments section. Summations over rectangular image regions can be computed much faster using integral images (summed-area-tables [2]). Integral images can be computed in linear time by dynamic programming, and once computed allow summations over rectangular regions (such as the one in (3)) to be computed in constant time.

4. EXPERIMENTS AND RESULTS

The image dataset that is used for evaluating the proposed methods was collected from our remote water quality monitoring system between May 2012 and June 2012. The data exhibited a wide variety of lighting and weather conditions, as well as a variety of ship types. A sample of the image data shown in Figure 4 demonstrates the complexity of the dataset. Approximately 3000 daylight color images with 640×480 pixels with different time intervals were annotated as the ground truth of the dataset. These images are separated into two categories: (1) no ship present and (2) ship present. For each group, 50 percent of the images were randomly chosen to build a model and the remainder of the images were subsequently used for testing.

To compare the performance and accuracy of the proposed methods, three background models are built: An Eigenbackgrounds model with 75 components and 60 bins, a GMM with 15 distributions and a GMM with 35 distributions. Figure 5 is a plot of the EVR of the Eigenbackgrounds model of our testing dataset. As can be seen from the plot, after 60 the ratio becomes steady and close to zero. This means the components after 60 carry very little background information. However, in order to gain the best performance 75 components are used for our experiments. Firstly, the histogram of local background residuals (HLBR) with different rectangular region sizes and histogram ranges are examined. Other image features including standard sum of error (SSE), a global pixel density histogram with 255 bins, local pixel density histograms over 2×2 and 4×4 sub-images were also tested for comparison. SSE is the sum of squared error between the back projected image and the original (which equals the square distance of the image from the PCA subspace). Classification[1] was carried out using a Radial Basis Function (RBF) kernel as it can handle a nonlinear decision boundary. The default parameters were used for the RBF kernel. All the experiments are carried out on a standard workstation with Intel Core i7 2600 3.4GHz CPU, and 16G RAM. Figure 6 shows model training results using 10 fold cross validation. It shows that the Eigenbackgrounds model performs better than GMM models.

Figure 4: illustraction of image data

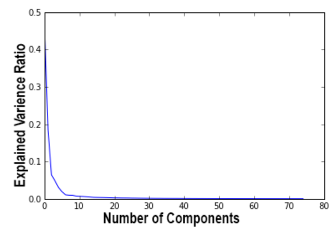

Figure 5: EVR of Eigenbackgrounds model

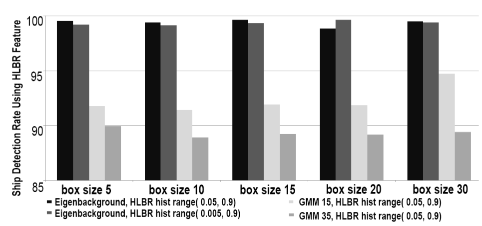

Figure 6: Models cross validation results

Figure 7 shows the ship detection rate using the testing image set. It can be seen that the GMM model with 15 distributions performs better than the GMM model with 35 distributions but poorer than the performance of the Eigenbackgrounds model. HLBR feature with rectangular region size 15 and a histogram range between 0.05 and 0.9 using an Eigenbackgrounds model achieves the best detection rate. However, the accuracy starts falling with both an increase

[1]Classification is performed in the Weka data analysis environment using LIBSVM implementation

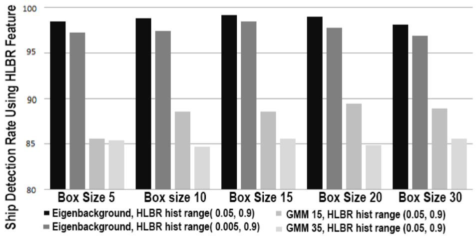

Figure 7: Ship detection results using HLBR feature

and decrease of the region size. Also, a histogram with a range of 0.05 to 0.9 performs better than a range of 0.005 and 0.9 but the variance is very small. Further experiments will be carried out in order to determine the best HLBR feature parameters.

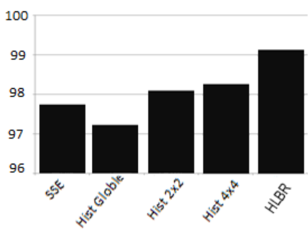

Figure 8: Comparison of different features based on Eigenbackgrounds.

Results in Figure 8 show that the HLBR feature achieves better performance than other features. This is because the HLBR feature calculates not just the distribution of foreground pixel values globally but also locally within small regions. Thus, it performs better when dealing with large amounts of noise.

Finally, it only took 24 seconds to build an Eigenbackgrounds model, but it took over 3 hours to build a GMM model. Each pixel in a GMM model is independent, therefore multi threading and process distribution can be used to reduce the time costs. However building the Eigenbackgrounds model is still much faster.

5. CONCLUSION

In this paper, we developed and evaluated an algorithm for automatically detecting the presence of ships in images captured by a very low frame rate remote visual sensing system. This follows an establishment of the importance of monitoring ship traffic at the site from an environmental monitoring perspective. Ships can be detected with a very high accuracy. This means that a visual sensor may be employed as a low cost platform for estimating conditions at the site in the future and may complement and enhance the use of a sophisticated in-situ sensor. The visual sensing platform may provide contextual information to increase the efficiency and effectiveness of the in-situ sensors creating a smarter multi-modal marine monitoring network.

6. ACKNOWLEDGEMENTS

QUESTOR, MESTECH, CLARITY

7. REFERENCES

[1] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz. Robust real-time unusual event detection using multiple fixed-location monitors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(3):555 –560, march 2008.

[2] F. C. Crow. Summed-area tables for texture mapping. *SIGGRAPH Comput. Graph.*, 18(3):207–212, Jan. 1984.

[3] M. Davidson, M. V. Koningsveld, A. de Kruif, J. Rawson, R. Holman, A. Lamberti, R. Medina, A. Kroon, and S. Aarninkhof. The CoastView project: Developing video-derived Coastal State Indicators in support of coastal zone management. *Coastal Engineering*, 54(6-7):463 – 475, 2007.

[4] E. A. Graham, E. M. Yuen, G. F. Robertson, W. J. Kaiser, M. P. Hamilton, and P. W. Rundel. Budburst and leaf area expansion measured with a novel mobile camera system and simple color thresholding. *Environmental and Experimental Botany*, 65(2-3):238 – 244, 2009.

[5] B. Lo and S. Velastin. Automatic congestion detection system for underground platforms. In *Intelligent Multimedia, Video and Speech Processing, 2001. Proceedings of 2001 International Symposium on*, pages 158 –161, 2001.

[6] C. Newcombe and D. MacDonald. Effects of suspended sediments on aquatic ecosystems. *North American Journal of Fisheries Management*, 11(1):72–82, 1991.

[7] E. O'Connor, A. Smeaton, and N. O'Connor. A multi-modal event detection system for river and coastal marine monitoring applications. In *OCEANS, 2011 IEEE - Spain*, pages 1 –10, june 2011.

[8] E. O'Connor, A. F. Smeaton, N. E. O'Connor, and F. Regan. A neural network approach to smarter sensor networks for water quality monitoring. *Sensors*, 12(4):4605–4632, 2012.

[9] M. Piccardi. Background subtraction techniques: a review. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 4, pages 3099 – 3104 vol.4, oct. 2004.

[10] A. Richardson, J. Jenkins, B. Braswell, D. Hollinger, S. Ollinger, , and M. Smith. Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. *Ecosystem Ecology*, 152(2):323–334, 2007.

[11] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2, pages 2 vol. (xxiii+637+663), 1999.

[12] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: real-time tracking of the human body. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):780 –785, jul 1997.